

Blended Latent Diffusion

Year: 2022 | Citations: 474 | Authors: Omri Avrahami, Ohad Fried, Dani Lischinski

Abstract

The tremendous progress in neural image generation, coupled with the emergence of seemingly omnipotent vision-language models has finally enabled text-based interfaces for creating and editing images. Handling generic images requires a diverse underlying generative model, hence the latest works utilize diffusion models, which were shown to surpass GANs in terms of diversity. One major drawback of diffusion models, however, is their relatively slow inference time. In this paper, we present an accelerated solution to the task of local text-driven editing of generic images, where the desired edits are confined to a user-provided mask. Our solution leverages a text-to-image Latent Diffusion Model (LDM), which speeds up diffusion by operating in a lower-dimensional latent space and eliminating the need for resource-intensive CLIP gradient calculations at each diffusion step. We first enable LDM to perform local image edits by blending the latents at each step, similarly to Blended Diffusion. Next we propose an optimization-based solution for the inherent inability of LDM to accurately reconstruct images. Finally, we address the scenario of performing local edits using thin masks. We evaluate our method against the available baselines both qualitatively and quantitatively and demonstrate that in addition to being faster, it produces more precise results.